Secrecy Degrees of Freedom of MIMO Broadcast Channels with Delayed CSIT

Sheng Yang, *Member, IEEE,* Mari Kobayashi, *Member, IEEE,*
Pablo Piantanida, *Member, IEEE,* Shlomo Shamai (Shitz), *Fellow, IEEE*

Abstract

The degrees of freedom (DoF) of the two-user Gaussian multiple-input and multiple-output (MIMO) broadcast channel with confidential message (BCC) is studied under the assumption that delayed channel state information (CSI) is available at the transmitter. We characterize the optimal secrecy DoF (SDoF) region and show that it can be achieved by a simple artificial noise alignment (ANA) scheme. The proposed scheme sends the confidential messages superposed with the artificial noise over several time slots. Exploiting delayed CSI, the transmitter aligns the transmit signal in such a way that the useful message can be extracted at the intended receiver but is completely drowned by the artificial noise at the unintended receiver. The proposed scheme can be interpreted as a non-trivial extension of Maddah-Ali Tse (MAT) scheme and enables us to quantify the resource overhead, or equivalently the DoF loss, to be paid for the secrecy communications.

Manuscript submitted to IEEE Transactions on Information Theory in December 2011.

S. Yang, M. Kobayashi, and P. Piantanida are with the Telecommunications department of SUPELEC, 3 rue Joliot-Curie, 91190 Gif-sur-Yvette, France. (e-mail: {sheng.yang, mari.kobayashi, pablo.piantanida}@supelec.fr)

S. Shamai (Shitz) is with Technion-Israel Institute of Technology, Haifa, Israel. (e-mail: sshlomo@ee.technion.ac.il)

This work was partially supported by the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

I. INTRODUCTION

We consider the two-user Gaussian multi-input multi-output broadcast channel with confidential messages (MIMO-BCC), where the transmitter sends two confidential messages to receivers A and B, respectively, while keeping each of them secret to the unintended receiver. By letting m, n_{A}, and n_{B} denote the number of antennas at the transmitter, receiver A, and receiver B, respectively, the corresponding channel outputs are given by

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{e}_t, \tag{1a}$$

$$\boldsymbol{z}_t = \boldsymbol{G}_t \boldsymbol{x}_t + \boldsymbol{b}_t, \quad t = 1, 2, \ldots, n, \tag{1b}$$

where $(\boldsymbol{y}_t, \boldsymbol{z}_t)$ denotes the observations at the receiver A and B, respectively, at time instant t; $\boldsymbol{H}_t \in \mathcal{H} \subseteq \mathbb{C}^{n_{\mathsf{A}} \times m}, \boldsymbol{G}_t \in \mathcal{G} \subseteq \mathbb{C}^{n_{\mathsf{B}} \times m}$ are the associated channel matrices; $(\boldsymbol{e}_t, \boldsymbol{b}_t)$ are assumed to be independent and identically distributed (i.i.d.) additive white Gaussian noises $\sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$; the input vector $\boldsymbol{x}_t \in \mathbb{C}^{m \times 1}$ is subject to the average power constraint

$$\frac{1}{n} \sum_{t=1}^{n} \mathrm{tr}(\boldsymbol{x}_t \boldsymbol{x}_t^{\mathsf{H}}) \leq P. \tag{2}$$

Furthermore, as in [1], we assume any arbitrary stationary fading process such that $\boldsymbol{H}_t$ and $\boldsymbol{G}_t$ are mutually independent and change from an instant to another one in an independent manner. Note that the channel at hand boils down to the conventional Gaussian MIMO wiretap channel where the transmitter wishes to send one message to the intended receiver while keeping it secret to the other one, namely, the eavesdropper.

The secrecy capacity region of the two-user MIMO Gaussian BCC with perfect channel state information at transmitter (CSIT)and receivers has been characterized in [2] (see also references therein). As a special case, the Gaussian MIMO wiretap channel has been extensively studied in [3]–[7]. However, the secrecy capacity of the MIMO Gaussian wiretap channel with general (imperfect) CSI at the transmitter remains open. Since a complete characterization of the capacity region in this case is very difficult (if not impossible), a number of contributions have focused on the so-called secrecy degrees of freedom (SDoF), by capturing the behavior in high signal-to-noise (SNR) regime (see [8]–[11] and references therein). References [8]–[10] investigated the compound models where channel uncertainty at the encoder is modeled as a set of finite channel states, while [11] investigated the scenario where the transmitter knows some temporal structure of the block-fading processes. A fundamental observation is that unless two channels enjoy asymmetric statistical properties[1], the perfect secrecy cannot be guaranteed under

[1]This may be in terms of asynchronous fading variation, different fading speed, number of antennas, etc..

a general CSIT assumption. In other words, if the statistics of the underlying channels seen by both receivers are symmetrical, additional side information (not necessarily instantaneous CSIT) is essential to ensure a positive SDoF, by introducing some asymmetry at the encoder. As a matter of fact, this reveals one of the major limitations of the wiretap model whose performance strongly depends on the quality of the channel state information at the transmitter side. Evidently, theoretically addressing CSI issues is of fundamental impact for secrecy systems.

Recently, in the context of multi-antenna broadcast channel, the pioneering work [1] showed that completely outdated channel state information at the transmitter is still very useful and increases the degrees of freedom of the multi-user channel. Motivated by this exciting result, the new assumption, commonly referred to as delayed CSIT, has since been applied to several multi-user settings, including the MIMO broadcast channel, X channel, and interference channel [12]–[15]. Non-trivial gain of degrees of freedom have been shown in all these settings with delayed CSIT. The main idea behind the utility of delayed CSIT can be best described with the term "retrospective interference alignment" introduced in [13] and [16]. That is, the knowledge of causal channel state is used to align the interference between users into a spatial/temporal subspace with a reduced dimension at each receiver.

In this paper, we study the impact of delayed CSIT on the secrecy degrees of freedom in a MIMO broadcast channel. In our setting, delayed CSI of a given receiver is available both at the transmitter and the other receiver[2], whereas each receiver knows its own instantaneous channel. Such a scenario is of practical interest since the receivers may send their channel states to the transmitter via delayed feedback links that may be overheard by the other receivers. We first characterize the optimal SDoF of the Gaussian MIMO wiretap channel with delayed CSIT. It is shown that delayed CSIT can significantly improve the SDoF, provided that the number of transmit antennas is larger than that of receive antennas, i.e., $m > \max(n_{\mathsf{A}}, n_{\mathsf{B}})$. In this case, we prove that a simple artificial noise alignment (ANA) scheme achieves the optimal SDoF. The proposed scheme sends the confidential symbols embedded by the artificial noise in such a way that the artificial noise is aligned in a subspace at the legitimate receiver while it fills the full signal space at the eavesdropper. The case of partial knowledge where the transmitter has delayed CSI only on the legitimate channel is also investigated. In this case, we show that a strictly smaller SDoF is achieved compared to the case with delayed CSIT on both channels. Then, we consider the two-user Gaussian MIMO-BCC and characterize the optimal SDoF region. The achievability follows

[2]Unless it is explicitly mentioned, we assume that delayed CSI of both channels is available to the transmitter, i.e., it observes $\boldsymbol{H}^{t-1}$ and $\boldsymbol{G}^{t-1}$ for every $t = 1, 2, \ldots$.

from an artificial noise alignment scheme adapted to convey two confidential messages. The proposed scheme can be seen as a non-trivial extension of the Maddah-Ali Tse (MAT) scheme. A simple comparison with the MAT scheme enables us to quantify the resource overhead, or equivalently the DoF loss, to be paid to guarantee the confidentiality of messages. Although delayed CSIT is found beneficial for a large range of transmit antennas analogy to the conclusions drawn for other network systems without secrecy constraints [1], [13], [14], we remark that the lack of perfect CSIT significantly degrades the performance of the secrecy systems.

The rest of the paper is organized as follows. Section II introduces the assumptions and some useful lemmas while Section III summarizes our main results on the optimal SDoF. Sections IV and V are devoted to proof of the main theorems. Finally, the paper is concluded in Section VI with some open problems and future perspectives.

II. Notations, Definitions, and Assumptions

A. Notations

Boldface lower-case letters $\boldsymbol{v}$ and upper-case letters $\boldsymbol{M}$ are used to denote vectors and matrices, respectively. We use the superscript notation X^n to denote a sequence $(X_1, \ldots, X_n)$ for any type of variables. Matrix transpose, Hermitian transpose, inverse, trace, and determinant are denoted by $\boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}^{\mathsf{H}}$, $\boldsymbol{A}^{-1}$, $\mathrm{tr}(\boldsymbol{A})$, and $\det(\boldsymbol{A})$, respectively. We let $\mathrm{diag}(\{\boldsymbol{A}_t\}_t)$ denote the block diagonal matrix with the matrices $\boldsymbol{A}_t$ as diagonal elements. Logarithm is in base 2 unless otherwise is specified. The differential entropy of X is denoted by $h(X)$. $(x)^+$ means $\max\{0, x\}$. The little-o notation $\mathsf{o}(\log P)$ stands for any real-valued function $f(P)$ such that $\lim_{P\to\infty} \frac{f(P)}{\log P} = 0$. The dot equality means the equality on the "pre-log" factor, i.e., $f(P) \doteq g(P)$ is equivalent to $f(P) = g(P) + \mathsf{o}(\log P)$; the dot inequalities $\dot{\geq}$ and $\dot{\leq}$ are similarly defined.

B. Assumptions and Definitions

The following assumptions and definitions will be applied in the rest of the paper.

Definition 1 (channel states): The channel matrices $\boldsymbol{H}_t$ and $\boldsymbol{G}_t$ are called the states of the channel at instant t. For simplicity, we also define the state matrix $\boldsymbol{S}_t$ as $\boldsymbol{S}_t = \begin{bmatrix} \boldsymbol{H}_t \\ \boldsymbol{G}_t \end{bmatrix}$.

Assumption 2.1 (delayed CSIT): At each time t, the states of the past $\boldsymbol{S}^{t-1}$ are known to all terminals. However, the instantaneous states $\boldsymbol{H}_t$ and $\boldsymbol{G}_t$ are only known to the respective receivers.

Under these assumptions, we define the code and the optimal SDoF region summarized below.

Definition 2 (code and SDoF region): A code for the Gaussian MIMO-BCC with delayed CSIT consists of:

- A sequence of stochastic encoders given by

$$\{F_t : \mathcal{W}_{\mathsf{A}} \times \mathcal{W}_{\mathsf{B}} \times \mathcal{H}^{t-1} \times \mathcal{G}^{t-1} \longmapsto \mathbb{C}^m\}_{t=1}^n,$$

  where the messages W_{A} and W_{B} are uniformly distributed over $\mathcal{W}_{\mathsf{A}}$ and $\mathcal{W}_{\mathsf{B}}$, respectively.
- The decoder A is given by the mapping $\hat{W}_{\mathsf{A}} : \mathbb{C}^{n_{\mathsf{A}} \times n} \times \mathcal{H}^n \times \mathcal{G}^{n-1} \longmapsto \mathcal{W}_{\mathsf{A}}$.
- The decoder B is given by the mapping $\hat{W}_{\mathsf{B}} : \mathbb{C}^{n_{\mathsf{B}} \times n} \times \mathcal{H}^{n-1} \times \mathcal{G}^{n} \longmapsto \mathcal{W}_{\mathsf{B}}$.

A SDoF pair $(d_{\mathsf{A}}, d_{\mathsf{B}})$ is said *achievable* if there exists a code that satisfies the reliability conditions at both receivers

$$\lim_{P\to\infty} \liminf_{n\to\infty} \frac{\log|\mathcal{W}_{\mathsf{A}}(n,P)|}{n\log P} \geq d_{\mathsf{A}}, \quad \lim_{P\to\infty} \limsup_{n\to\infty} \Pr\left\{W_{\mathsf{A}} \neq \hat{W}_{\mathsf{A}}\right\} = 0, \tag{3}$$

$$\lim_{P\to\infty} \liminf_{n\to\infty} \frac{\log|\mathcal{W}_{\mathsf{B}}(n,P)|}{n\log P} \geq d_{\mathsf{B}}, \quad \lim_{P\to\infty} \limsup_{n\to\infty} \Pr\left\{W_{\mathsf{B}} \neq \hat{W}_{\mathsf{B}}\right\} = 0, \tag{4}$$

as well as the perfect secrecy condition

$$\lim_{P\to\infty} \limsup_{n\to\infty} \frac{I(W_{\mathsf{A}}; \boldsymbol{z}^n, \boldsymbol{S}^n)}{n\log P} = 0, \tag{5}$$

$$\lim_{P\to\infty} \limsup_{n\to\infty} \frac{I(W_{\mathsf{B}}; \boldsymbol{y}^n, \boldsymbol{S}^n)}{n\log P} = 0. \tag{6}$$

The union of all achievable pairs $(d_{\mathsf{A}}, d_{\mathsf{B}})$ is called the optimal SDoF region.

Assumption 2.2 (channel symmetry): At any instant t, the rows of the state matrix $\boldsymbol{S}_t$ are independent and identically distributed. Furthermore, we limit ourselves to the class of fading processes in which the state matrix $\boldsymbol{S}_t$ has full rank $\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}$ almost surely at any time instant t.[3]

As direct consequences of the channel symmetry, we readily have that the marginal distributions of any antenna output are equal conditioned on the same previous observations and/or the source message. Namely, we have the following property.

Property 2.1 (channel output symmetry): Let $\Omega_t = \{y_{1,t}, \ldots, y_{n_{\mathsf{A}},t}, z_{1,t}, \ldots, z_{n_{\mathsf{B}},t}\}$ be the collection of random variables representing all antenna outputs at time instant t. Then, for any subset $\omega_{\mathcal{J}}$ and $\omega_{\mathcal{K}}$ of random variables in Ω_t satisfying $|\omega_{\mathcal{J}}| = |\omega_{\mathcal{K}}|$, we have

$$\Pr(\omega_{\mathcal{J}} \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}, U_t) = \Pr(\omega_{\mathcal{K}} \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}, U_t) \tag{7}$$

for any random variables $U_t \leftrightarrow (\boldsymbol{H}_t, \boldsymbol{G}_t, W) \leftrightarrow \Omega_t$ with $t = \{1, \ldots, n\}$ that form a Markov chain.

[3]This assumption is used to prove the achievability although the converse proof does not need such an assumption.

Using the fact that current channel outputs do not depend on the future channel realizations, we can easily show that Property 2.1 also holds when we add the conditioning on $\boldsymbol{S}^n$, namely,

$$h(\omega_{\mathcal{J}} \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}, \boldsymbol{S}^n, W) = h(\omega_{\mathcal{K}} \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}, \boldsymbol{S}^n, W). \tag{8}$$

In the following, we omit the conditioning on $\boldsymbol{S}^n$ for notation simplicity.

C. Preliminaries

For sake of clarity, we collect the results that will be used repeatedly in the rest of the paper. First, the following lemma is the direct consequences of the channel output symmetry.

Lemma 1 (properties of channel symmetry): The following inequalities hold under the *channel output symmetry* Property 2.1:

$$\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}\, h(\boldsymbol{z}^n) \,\dot{\geq}\, n_{\mathsf{B}}\, h(\boldsymbol{y}^n, \boldsymbol{z}^n), \tag{9a}$$

$$\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}\, h(\boldsymbol{y}^n) \,\dot{\geq}\, n_{\mathsf{A}}\, h(\boldsymbol{y}^n, \boldsymbol{z}^n), \tag{9b}$$

$$\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}\, h(\boldsymbol{z}^n) \,\dot{\geq}\, n_{\mathsf{B}}\, h(\boldsymbol{y}^n), \tag{9c}$$

$$\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}\, h(\boldsymbol{y}^n) \,\dot{\geq}\, n_{\mathsf{A}}\, h(\boldsymbol{z}^n). \tag{9d}$$

Furthermore, same inequalities hold true conditional on W.

Proof: The first two inequalities are proved in Appendix A. To prove (9c), from (9a), we have

$$h(\boldsymbol{z}^n) \,\dot{\geq}\, \frac{n_{\mathsf{B}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} h(\boldsymbol{y}^n, \boldsymbol{z}^n) \tag{10}$$

$$\geq \frac{n_{\mathsf{B}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} h(\boldsymbol{y}^n), \tag{11}$$

where the last inequality comes from the fact that $h(\boldsymbol{z}^n \,|\, \boldsymbol{y}^n) \geq h(\boldsymbol{z}^n \,|\, \boldsymbol{y}^n, \boldsymbol{x}^n) = h(\boldsymbol{b}^n) = \mathsf{o}(\log P)$. Same steps can be applied to obtain (9d). ■

Then, all the achievable DoF results are essentially based on the rank of the channel matrices.

Lemma 2: For any matrix $\boldsymbol{A}$ which does not depend on P, we have

$$\lim_{P\to\infty} \frac{\log\det(\mathbf{I} + P\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}})}{\log P} = \mathrm{rank}(\boldsymbol{A}). \tag{12}$$

Proof: Let $(\sigma_1, \ldots, \sigma_r)$ be the $r \triangleq \mathrm{rank}(\boldsymbol{A})$ non-zero singular values of $\boldsymbol{A}$. Then, we have that

$$\log\det(\mathbf{I} + P\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}}) = \sum_{k=1}^{r} \log(1 + P\sigma_k^2) \doteq r \log P,$$

since the non-zero singular values do not depend on P and thus do not vanish with P either. ■

III. Main Results

In this section, we highlight our main results on the optimal SDoF of the Gaussian MIMO wiretap channel and then on the more general Gaussian MIMO broadcast channel with confidential messages. We shall interpret the results through comparisons and numerical evaluations.

A. Wiretap Channel

Theorem 1 (wiretap channel with delayed CSIT): In presence of delayed CSIT on both the legitimate channel and the eavesdropper channel, the optimal SDoF of the Gaussian MIMO wiretap channel with $m, n_{\mathsf{A}}, n_{\mathsf{B}}$ antennas at the transmitter, the legitimate receiver, the eavesdropper, respectively, is given by

$$d_s(n_{\mathsf{A}}, n_{\mathsf{B}}, m) = \begin{cases} 0, & m \leq n_{\mathsf{B}}, \\ m - n_{\mathsf{B}}, & n_{\mathsf{B}} < m \leq n_{\mathsf{A}}, \\ \dfrac{n_{\mathsf{A}} m (m - n_{\mathsf{B}})}{n_{\mathsf{A}} n_{\mathsf{B}} + m(m - n_{\mathsf{B}})}, & \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}, \\ \dfrac{n_{\mathsf{A}}(n_{\mathsf{A}} + n_{\mathsf{B}})}{n_{\mathsf{A}} + 2n_{\mathsf{B}}}, & m > n_{\mathsf{A}} + n_{\mathsf{B}}. \end{cases} \tag{13}$$

In the wiretap setting, it is not always reasonable to assume any CSI on the eavesdropper channel at the transmitter side. In this case, we may consider delayed CSIT only on the legitimate channel and without CSIT on the eavesdropper channel. With this asymmetric CSI assumption, hereafter referred to as delayed partial CSIT, we can show that a strictly positive yet smaller SDoF than delayed CSIT on both channels is still achievable for a wide range of number of antennas.

Theorem 2 (wiretap channel with delayed partial CSIT): In presence of delayed partial CSIT, either on the legitimate channel or the eavesdropper channel, the following SDoF is achievable for MIMO Gaussian wiretap channel for $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$

$$d_s^{\text{partial}}(n_{\mathsf{A}}, n_{\mathsf{B}}, m) = \begin{cases} \dfrac{n_{\mathsf{A}}(m - n_{\mathsf{B}})}{m}, & \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}, \\ \dfrac{n_{\mathsf{A}}^2}{n_{\mathsf{A}} + n_{\mathsf{B}}}, & m > n_{\mathsf{A}} + n_{\mathsf{B}}. \end{cases} \tag{14}$$

Note that it is the best known achievable SDoF in this setting, although the converse is yet to be proved.

In order to quantify the benefit of delayed CSIT, we summarize the SDoF with perfect, delayed and without CSIT in Table I and provide an example with $n_{\mathsf{A}} = 3, n_{\mathsf{B}} = 2$ in Fig. 1. We remark that delayed CSIT is beneficial only when the number of transmit antennas is larger than the number of receive antennas, i.e., $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$, since the SDoF is $(m - n_{\mathsf{B}})^+$ for $m \leq \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$ with perfect,

delayed, and without CSIT. As the number m of transmit antennas increases, the SDoF grows until $m = n_{\mathsf{A}} + n_{\mathsf{B}}$ for perfect and delayed CSIT while it does not increase with m without CSIT. It appears that with both perfect and delayed CSIT, we cannot exploit any gain for m beyond $n_{\mathsf{A}}+n_{\mathsf{B}}$. Furthermore, we remark that delayed CSIT only on the legitimate channel incurs a non-negligible loss compared to delayed CSIT on both channels. This is because the transmitter without CSI on the eavesdropper channel cannot access to the signal overheard by the eavesdropper, which reduces the signal dimension to be exploited by the legitimate receiver.

TABLE I

COMPARISON OF THE SDOF UNDER DIFFERENT CSIT ASSUMPTIONS FOR $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$.

CSIT	$\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m < n_{\mathsf{A}} + n_{\mathsf{B}}$	$m \geq n_{\mathsf{A}} + n_{\mathsf{B}}$
perfect	$m - n_{\mathsf{B}}$	n_{A}
delayed	$\frac{n_{\mathsf{A}} m(m-n_{\mathsf{B}})}{n_{\mathsf{A}} n_{\mathsf{B}} + m(m-n_{\mathsf{B}})}$	$\frac{n_{\mathsf{A}}(n_{\mathsf{A}}+n_{\mathsf{B}})}{n_{\mathsf{A}}+2n_{\mathsf{B}}}$
delayed partial	$\frac{n_{\mathsf{A}}(m-n_{\mathsf{B}})}{m}$	$\frac{n_{\mathsf{A}}^2}{n_{\mathsf{A}}+n_{\mathsf{B}}}$
no	$(n_{\mathsf{A}} - n_{\mathsf{B}})^+$	$(n_{\mathsf{A}} - n_{\mathsf{B}})^+$

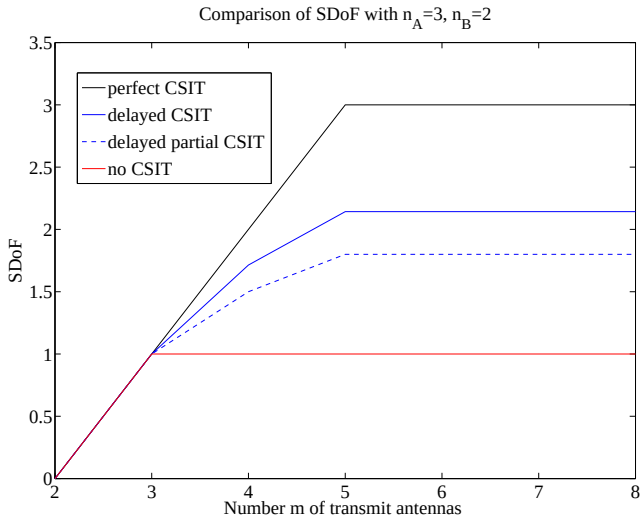

Fig. 1. SDoF with $n_{\mathsf{A}} = 3$ and $n_{\mathsf{B}} = 2$ with perfect, delayed, and no CSIT.

B. Broadcast channel with confidential messages

Next, we present the achievable SDoF region of the two-user MIMO-BCC with delayed CSIT.

Theorem 3 (BCC with delayed CSIT): The optimal SDoF region $\mathcal{R}_{\text{BCC}}$ of the two-user MIMO-BCC with delayed CSIT is given as a set of non-negative $(d_{\mathsf{A}}, d_{\mathsf{B}})$ satisfying

$$\frac{d_{\mathsf{A}}}{d_s(n_{\mathsf{A}}, n_{\mathsf{B}}, m)} + \frac{d_{\mathsf{B}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} \leq 1, \tag{15a}$$

$$\frac{d_{\mathsf{A}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} + \frac{d_{\mathsf{B}}}{d_s(n_{\mathsf{B}}, n_{\mathsf{A}}, m)} \leq 1, \tag{15b}$$

for any $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$. If $n_{\mathsf{B}} < m \leq n_{\mathsf{A}}$, we have $d_{\mathsf{A}} \leq m - n_{\mathsf{B}}$ and $d_{\mathsf{B}} = 0$, whereas if $n_{\mathsf{A}} < m \leq n_{\mathsf{B}}$, we have $d_{\mathsf{A}} = 0$ and $d_{\mathsf{B}} \leq m - n_{\mathsf{A}}$.

Corollary 3.1: For the case $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$, the SDoF region is characterized by the two corner points $(0, d_s(n_{\mathsf{B}}, n_{\mathsf{A}}, m))$, $(d_s(n_{\mathsf{A}}, n_{\mathsf{B}}, m), 0)$ and the sum SDoF point given by

$$(d_{\mathsf{A}}, d_{\mathsf{B}}) = \begin{cases} \left(\dfrac{n_{\mathsf{A}}(m - n_{\mathsf{B}})}{m}, \dfrac{n_{\mathsf{B}}(m - n_{\mathsf{A}})}{m}\right), & \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}} \\ \left(\dfrac{n_{\mathsf{A}}^2}{n_{\mathsf{A}} + n_{\mathsf{B}}}, \dfrac{n_{\mathsf{B}}^2}{n_{\mathsf{A}} + n_{\mathsf{B}}}\right), & m > n_{\mathsf{A}} + n_{\mathsf{B}}. \end{cases} \tag{16}$$

Remark 3.1: We can find trivial upper bounds to the above SDoF region for the case of $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$. On one hand, the SDoF region with delayed CSIT is dominated by the SDoF region with perfect CSIT. The SDoF region with perfect CSIT is square connecting three corner points $(\min\{n_{\mathsf{A}}, m - n_{\mathsf{B}}\}, 0)$, $(\min\{n_{\mathsf{A}}, m - n_{\mathsf{B}}\}, \min\{m - n_{\mathsf{A}}, n_{\mathsf{B}}\})$, and $(0, \min\{m - n_{\mathsf{A}}, n_{\mathsf{B}}\})$. We can also compare the above SDoF region with delayed CSIT and the DoF region of the two-user MIMO-BC with delayed constraint [12], given by

$$\frac{d_{\mathsf{A}}}{\min\{m, n_{\mathsf{A}}\}} + \frac{d_{\mathsf{B}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} \leq 1, \tag{17a}$$

$$\frac{d_{\mathsf{A}}}{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}} + \frac{d_{\mathsf{B}}}{\min\{m, n_{\mathsf{B}}\}} \leq 1. \tag{17b}$$

Obviously, since SDoF is always upper bounded by DoF of the MIMO channel, namely $d_s(n_{\mathsf{A}}, n_{\mathsf{B}}, m) \leq \min\{n_{\mathsf{A}}, m\}$ and $d_s(n_{\mathsf{B}}, n_{\mathsf{A}}, m) \leq \min\{n_{\mathsf{B}}, m\}$, the SDoF region is dominated by the DoF region.

We provide an insight to the proposed artificial noise alignment scheme which achieves the sum SDoF point $\left(\frac{1}{2}, \frac{1}{2}\right)$ over the two-user MISO-BCC. Let us consider the four-slot scheme where the transmitter sends six independent Gaussian distributed symbols $\boldsymbol{u} \triangleq [u_1 \ u_2]^{\mathsf{T}}$, $\boldsymbol{v}_{\mathsf{A}} \triangleq [v_{11} \ v_{12}]^{\mathsf{T}}$, $\boldsymbol{v}_{\mathsf{B}} \triangleq [v_{21} \ v_{22}]^{\mathsf{T}}$ whose powers scale equally with P. Specifically, the transmit vectors are given by

$$\boldsymbol{x}_1 = \boldsymbol{u}, \ \ \boldsymbol{x}_2 = \boldsymbol{v}_{\mathsf{A}} + \begin{bmatrix} \boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{u} \\ 0 \end{bmatrix}, \boldsymbol{x}_3 = \boldsymbol{v}_{\mathsf{B}} + \begin{bmatrix} \boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{u} \\ 0 \end{bmatrix}, \boldsymbol{x}_4 = \begin{bmatrix} (\boldsymbol{g}_2^{\mathsf{T}}\boldsymbol{v}_{\mathsf{A}} + g_{21}\boldsymbol{h}_1^{\mathsf{T}}\boldsymbol{u}) + (\boldsymbol{h}_3^{\mathsf{T}}\boldsymbol{v}_{\mathsf{B}} + h_{31}\boldsymbol{g}_1^{\mathsf{T}}\boldsymbol{u}) \\ 0 \end{bmatrix}, \tag{18}$$

where, for simplicity of demonstration, we omit the scaling factors that fulfill the power constraint (2). Note that this simplification, also adopted in [1] and other related works, does not affect the high SNR

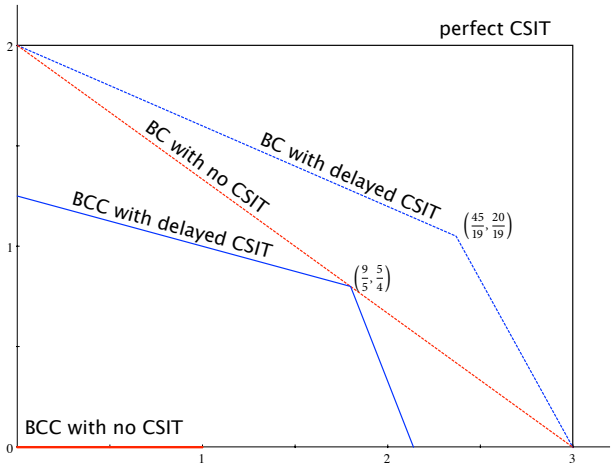

Fig. 2. The two-user DoF/SDoF region with $m = 5, n_{\mathsf{A}} = 3, n_{\mathsf{B}} = 2$.

analysis carried out here. The following remarks are in order. First, it can be easily shown that, at receiver A, $\boldsymbol{v}_{\mathsf{A}}$ lies in a two-dimensional subspace, while the unintended signal $\boldsymbol{v}_{\mathsf{B}}$ plus the artificial noise are aligned in another two-dimensional subspace. Thus, the intended message can be recovered through $\boldsymbol{v}_{\mathsf{A}}$ from the four-dimensional observation at receiver A. Second, $\boldsymbol{v}_{\mathsf{A}}$ is drowned in the observation at receiver B. More precisely, at receiver B, $\boldsymbol{v}_{\mathsf{A}}$ is squeezed into a one-dimension subspace filled with artificial noise, which makes it impossible to recover any useful information about A. Due to the symmetry, the same holds for $\boldsymbol{v}_{\mathsf{B}}$. Therefore, we can send simultaneously two confidential symbols to each receiver over four slots, yielding the sum SDoF point $\left(\frac{1}{2}, \frac{1}{2}\right)$.

The four-slot scheme contains two special cases of interest. If we consider the MISO wiretap channel where the transmitter wishes to convey $\boldsymbol{v}_{\mathsf{A}}$ to receiver A while keeping it secret to receiver B, we let $\boldsymbol{v}_{\mathsf{B}} = \mathbf{0}$ and ignore the third time slot. This provides a SDoF of $\frac{2}{3}$. If we consider the two-user MISO-BC without secrecy constraint, we remove the artificial noise transmission by letting $\boldsymbol{u} = \mathbf{0}$ and ignoring the first time slot. This boils down to the MAT scheme [1]. The four-slot scheme as well as the more general artificial noise alignment scheme presented in Section V is indeed a non-trivial extension of retrospective interference alignment schemes for MIMO broadcast channels [1], [12] to secure communications. The comparison with the three-slot MAT scheme can be interpreted as follows. The messages can be kept secret at a price of an additional resource (one slot), which appears as a DoF loss with respect to the communication systems without secrecy constraint.

In order to visualize the DoF loss due to the secrecy constraints, we provide an example of the achievable DoF/SDoF regions with $m = 5$, $n_{\mathsf{A}} = 3$, and $n_{\mathsf{B}} = 2$ in Fig. 2. For the case of perfect CSIT, the SDoF region and the DoF region are square. In the MIMO-BC, we send $(n_{\mathsf{A}}^2(n_{\mathsf{A}}+n_{\mathsf{B}}), n_{\mathsf{B}}^2(n_{\mathsf{A}}+n_{\mathsf{B}})) = (45, 20)$ private symbols to receiver A and B, respectively, over a duration of $n_{\mathsf{A}}^2 + n_{\mathsf{B}}^2 + n_{\mathsf{A}}n_{\mathsf{B}} = 19$ slots, yielding the DoF $\left(\frac{45}{19}, \frac{20}{19}\right)$, as shown in [12]. Under the perfect secrecy constraints, we need an extra phase of the artificial noise transmission of $n_{\mathsf{A}}n_{\mathsf{B}} = 6$ slots to convey two streams securely. This yields the SDoF of $\left(\frac{9}{5}, \frac{4}{5}\right)$. The comparison with the DoF region of the MIMO-BC can be interpreted in either an optimistic or a pessimistic way. On one hand, the benefit of delayed CSIT is more significant for the SDoF region. On the other hand, we also observe that the lack of accurate CSIT decreases substantially the SDoF, which implies that the secure communications are very sensitive to the quality of CSIT.

IV. Wiretap Channel: Proofs of Theorems 1 and 2

A. Converse proof of Theorem 1

We are now ready to provide the converse by considering different cases below.

1) Case $m \leq n_{\mathsf{B}}$: From Fano's inequality and the secrecy constraint, we have

$$
\begin{aligned}
n(R - \mathsf{o}(\log P)) &\leq I(W; \boldsymbol{y}^n) - I(W; \boldsymbol{z}^n) && (19)\\
&= I(W; \boldsymbol{y}^n \,|\, \boldsymbol{z}^n) - I(W; \boldsymbol{z}^n \,|\, \boldsymbol{y}^n) && (20)\\
&= h(\boldsymbol{y}^n \,|\, \boldsymbol{z}^n) - h(\boldsymbol{y}^n \,|\, \boldsymbol{z}^n, W) - I(W; \boldsymbol{z}^n \,|\, \boldsymbol{y}^n) && (21)\\
&\leq h(\boldsymbol{y}^n \,|\, \boldsymbol{z}^n) && (22)\\
&\leq \sum_{t=1}^{n} h(\boldsymbol{y}_t \,|\, \boldsymbol{z}_t) && (23)\\
&= \sum_{t=1}^{n} h(\boldsymbol{y}_t - \boldsymbol{H}_t\, \hat{\boldsymbol{x}}_{\boldsymbol{z},t} \,|\, \boldsymbol{z}_t) && (24)\\
&= \sum_{t=1}^{n} h(\boldsymbol{e}_t - \boldsymbol{H}_t\, (\hat{\boldsymbol{x}}_{\boldsymbol{z},t} - \boldsymbol{x}_t) \,|\, \boldsymbol{z}_t) && (25)\\
&= \sum_{t=1}^{n} h(\boldsymbol{e}_t - \boldsymbol{H}_t\, (\hat{\boldsymbol{x}}_{\boldsymbol{z},t} - \boldsymbol{x}_t)) && (26)\\
&= n\, \mathsf{o}(\log P), && (27)
\end{aligned}
$$

where (22) is from the fact that both $I(W; \boldsymbol{z}^n \,|\, \boldsymbol{y}^n)$ and $h(\boldsymbol{y}^n \,|\, \boldsymbol{z}^n, W)$ are non-negative; in (24) we use the fact that translations preserve differential entropy and let $\hat{\boldsymbol{x}}_{\boldsymbol{z},t}$ denote the MMSE estimation of $\boldsymbol{x}_t$ given $\boldsymbol{z}_t$; the last equality holds because the estimation error does not scale with P.

2) Case $n_B < m \leq \max\{n_A, n_B\}$: Since this case happens only when $n_{\mathsf{B}} < m \leq n_{\mathsf{A}}$, we can assume $m \leq n_{\mathsf{A}}$. Starting from (22), we have

$$n(R - \mathsf{o}(\log P)) \leq h(\boldsymbol{y}^n \,|\, \boldsymbol{z}^n) \tag{28}$$

$$\dot{\leq} \frac{\min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\} - n_{\mathsf{B}}}{n_{\mathsf{B}}} h(\boldsymbol{z}^n) \tag{29}$$

$$\leq n(m - n_{\mathsf{B}}) \log P + n\, \mathsf{o}(\log P), \tag{30}$$

where (29) follows straightforwardly from (9a); the last inequality comes from the fact that i.i.d. Gaussian variables maximize the differential entropies under the variance constraint.

3) Case $m > \max\{n_A, n_B\}$: In the following we let $\tilde{m} = \min\{m, n_{\mathsf{A}} + n_{\mathsf{B}}\}$ for notation simplicity. We remark that two upper bounds can be obtained as a direct consequence of Lemma 1. One one hand, (29) still holds

$$I(W; \boldsymbol{y}^n) - I(W; \boldsymbol{z}^n) \dot{\leq} \frac{\tilde{m} - n_{\mathsf{B}}}{n_{\mathsf{B}}} h(\boldsymbol{z}^n). \tag{31}$$

On the other hand, we have

$$I(W; \boldsymbol{y}^n) - I(W; \boldsymbol{z}^n) = h(\boldsymbol{y}^n) - h(\boldsymbol{y}^n \,|\, W) - h(\boldsymbol{z}^n) + h(\boldsymbol{z}^n \,|\, W) \tag{32}$$

$$\dot{\leq} h(\boldsymbol{y}^n) + \left(1 - \frac{n_{\mathsf{A}}}{\tilde{m}}\right) h(\boldsymbol{z}^n \,|\, W) - h(\boldsymbol{z}^n) \tag{33}$$

$$\leq h(\boldsymbol{y}^n) - \frac{n_{\mathsf{A}}}{\tilde{m}} h(\boldsymbol{z}^n), \tag{34}$$

where (33) follows from (9d); (34) follows from $h(\boldsymbol{z}^n \,|\, W) \leq h(\boldsymbol{z}^n)$; By combining the above two upper bounds, we readily have

$$n(R - \mathsf{o}(\log P)) \leq I(W; \boldsymbol{y}^n) - I(W; \boldsymbol{z}^n) \tag{35}$$

$$\dot{\leq} \min\left\{ \frac{\tilde{m} - n_{\mathsf{B}}}{n_{\mathsf{B}}} h(\boldsymbol{z}^n), h(\boldsymbol{y}^n) - \frac{n_{\mathsf{A}}}{\tilde{m}} h(\boldsymbol{z}^n) \right\} \tag{36}$$

$$\leq \max_{\alpha} \max_{\beta} \min\left\{ \frac{\tilde{m} - n_{\mathsf{B}}}{n_{\mathsf{B}}} \beta, \alpha - \frac{n_{\mathsf{A}}}{\tilde{m}} \beta \right\} \tag{37}$$

$$\leq \max_{\alpha} \alpha \left(1 + \frac{n_{\mathsf{A}} n_{\mathsf{B}}}{\tilde{m}(\tilde{m} - n_{\mathsf{B}})}\right)^{-1} \tag{38}$$

$$\dot{\leq} \left(1 + \frac{n_{\mathsf{A}} n_{\mathsf{B}}}{\tilde{m}(\tilde{m} - n_{\mathsf{B}})}\right)^{-1} n_{\mathsf{A}} n \log P, \tag{39}$$

where (37) is because the RHS of (36) can only increase by maximizing it over both entropies $\alpha = h(\boldsymbol{y}^n), \beta = h(\boldsymbol{z}^n)$; in (38) the inner maximization is solved by equalizing two terms inside $\min$, and finally we use $h(\boldsymbol{y}^n) \leq n_{\mathsf{A}} n \log P + \mathsf{o}(\log P)$. This establishes the converse proof.

B. Achievability proof of Theorem 1

In the following, we wish to show the achievability of the SDoF. As in the converse part, we consider separately the cases for different m. Note that only two ranges of m need to be considered. The first one is $n_{\mathsf{B}} \leq m \leq \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$ and the other one is $\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$. For $m < n_{\mathsf{B}}$, the SDoF is zero. For $m > n_{\mathsf{A}} + n_{\mathsf{B}}$, the converse shows that it is useless in terms of SDoF to set more than $n_{\mathsf{A}} + n_{\mathsf{B}}$ antennas.

1) Case $n_B \leq m \leq \max\{n_A, n_B\}$: For this case, we need to show that $d_s(n_{\mathsf{A}}, n_{\mathsf{B}}, m) = m - n_{\mathsf{B}}$ is achievable for $n_{\mathsf{B}} < m \leq n_{\mathsf{A}}$. This can be simply done by sending a vector of m symbols of which $m - n_{\mathsf{B}}$ symbols $\boldsymbol{v}$ are useful message and the other n_{B} symbols $\boldsymbol{u}$ are artificial noise (or a random message). The legitimate receiver can decode all m symbols and therefore extract the useful message, i.e.,

$$I(\boldsymbol{v};\boldsymbol{y}) = I(\boldsymbol{v},\boldsymbol{u};\boldsymbol{y}) - I(\boldsymbol{u};\boldsymbol{y}\,|\,\boldsymbol{v}) \tag{40}$$

$$= \log\det\left(\mathbf{I}_{n_{\mathsf{A}}} + \frac{P}{m}\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\right) - \log\det\left(\mathbf{I}_{n_{\mathsf{A}}} + \frac{n_{\mathsf{B}}P}{m}\boldsymbol{H}\begin{bmatrix}\mathbf{I}_{n_{\mathsf{B}}} \\ 0_{m-n_{\mathsf{B}}}\end{bmatrix}\boldsymbol{H}^{\mathsf{H}}\right) \tag{41}$$

$$\doteq (m - n_{\mathsf{B}})\log P, \tag{42}$$

while the eavesdropper channel is inflated by the random message and does not expose more than a vanishing fraction of the useful message, i.e.,

$$I(\boldsymbol{v};\boldsymbol{z}) = I(\boldsymbol{v},\boldsymbol{u};\boldsymbol{z}) - I(\boldsymbol{u};\boldsymbol{z}\,|\,\boldsymbol{v}) \tag{43}$$

$$= \log\det\left(\mathbf{I}_{n_{\mathsf{B}}} + \frac{P}{m}\boldsymbol{G}\boldsymbol{G}^{\mathsf{H}}\right) - \log\det\left(\mathbf{I}_{n_{\mathsf{B}}} + \frac{n_{\mathsf{B}}P}{m}\boldsymbol{G}\begin{bmatrix}\mathbf{I}_{n_{\mathsf{B}}} \\ 0_{m-n_{\mathsf{B}}}\end{bmatrix}\boldsymbol{G}^{\mathsf{H}}\right) \tag{44}$$

$$\doteq n_{\mathsf{B}}\log P - n_{\mathsf{B}}\log P \tag{45}$$

$$= 0, \tag{46}$$

where we used the fact that $\text{rank}(\boldsymbol{H}) = m$ and $\text{rank}(\boldsymbol{G}) = n_{\mathsf{B}}$. Note that (41) and (44) are obtained by applying independent Gaussian signaling to $\boldsymbol{v}$ and $\boldsymbol{u}$ with proper covariance corresponding to the power constraint. This assumption will be implicitly applied in the rest of the paper.

2) Case $\max\{n_A, n_B\} < m \leq n_A + n_B$: The proposed scheme combines the artificial noise with the Maddah-Ali Tse (MAT) alignment scheme [1]. The main idea of this scheme is to send the artificial noise such that it fills the eavesdropper's observation and hides the confidential message, while it shall be aligned in a reduced subspace at the legitimate receiver. The proposed three-phase scheme is presented in Table II, where the signal model without thermal noise is described concisely with the block matrix

TABLE II

PROPOSED THREE-PHASE SCHEME FOR $\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$.

Phase 1 ($t \in \mathcal{T}_1$)	Phase 2 ($t \in \mathcal{T}_2$)	Phase 3 ($t \in \mathcal{T}_3$)
$\boldsymbol{x}_1 = \boldsymbol{u}$	$\boldsymbol{x}_2 = \boldsymbol{v} + \boldsymbol{\Theta}\tilde{\boldsymbol{y}}_1$	$\boldsymbol{x}_3 = \boldsymbol{\Phi}\tilde{\boldsymbol{z}}_2$
$\tilde{\boldsymbol{y}}_1 = \tilde{\boldsymbol{H}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{y}}_2 = \tilde{\boldsymbol{H}}_2\boldsymbol{v} + \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{y}}_3 = \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{v} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\boldsymbol{u}$
$\tilde{\boldsymbol{z}}_1 = \tilde{\boldsymbol{G}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{z}}_2 = \tilde{\boldsymbol{G}}_2\boldsymbol{v} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{z}}_3 = \tilde{\boldsymbol{G}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{v} + \tilde{\boldsymbol{G}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\boldsymbol{u}$

TABLE III

LENGTH OF THREE PHASES $\{\tau_i\}$ FOR DIFFERENT $m, n_{\mathsf{A}}, n_{\mathsf{B}}$.

	$\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$	$m > n_{\mathsf{A}} + n_{\mathsf{B}}$
τ_1	$n_{\mathsf{A}} n_{\mathsf{B}}$	$n_{\mathsf{A}} n_{\mathsf{B}}$
τ_2	$n_{\mathsf{A}}(m - n_{\mathsf{B}})$	n_{A}^2
τ_3	$(m - n_{\mathsf{A}})(m - n_{\mathsf{B}})$	$n_{\mathsf{A}} n_{\mathsf{B}}$
total duration $\sum_{i=1}^{3} \tau_i$	$n_{\mathsf{A}} n_{\mathsf{B}} + m(m - n_{\mathsf{B}})$	$2n_{\mathsf{A}} n_{\mathsf{B}} + n_{\mathsf{A}}^2$

notation:

$$\boldsymbol{u} = \left[\begin{array}{ccc} \boldsymbol{u}_1^{\mathsf{T}} & \ldots & \boldsymbol{u}_{\tau_1}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}} \in \mathbb{C}^{m\tau_1 \times 1}, \qquad \boldsymbol{v} = \left[\begin{array}{ccc} \boldsymbol{v}_1^{\mathsf{T}} & \ldots & \boldsymbol{v}_{\tau_2}^{\mathsf{T}} \end{array} \right]^{\mathsf{T}} \in \mathbb{C}^{m\tau_2 \times 1}, \tag{47}$$

$$\tilde{\boldsymbol{H}}_i = \mathrm{diag}\left(\{\boldsymbol{H}_t\}_{t \in \mathcal{T}_i}\right) \in \mathbb{C}^{n_{\mathsf{A}}\tau_i \times m\tau_i}, \qquad \tilde{\boldsymbol{G}}_i = \mathrm{diag}\left(\{\boldsymbol{G}_t\}_{t \in \mathcal{T}_i}\right) \in \mathbb{C}^{n_{\mathsf{B}}\tau_i \times m\tau_i}, \tag{48}$$

$$\boldsymbol{\Theta} \in \mathbb{C}^{m\tau_2 \times n_{\mathsf{A}}\tau_1}, \qquad \boldsymbol{\Phi} \in \mathbb{C}^{m\tau_3 \times n_{\mathsf{B}}\tau_2}, \tag{49}$$

where τ_i denotes the length of phase i for $i = 1$, 2, and 3 given in Table III.

The three phases are explained as follows:

- *Phase 1,* $t \in \mathcal{T}_1 \triangleq \{1, \ldots, \tau_1\}$*:* ***sending the artificial noise***. The $m\tau_1$ symbols sent in τ_1 time slots is represented by $\boldsymbol{u}$.
- *Phase 2,* $t \in \mathcal{T}_2 \triangleq \{\tau_1 + 1, \ldots, \tau_1 + \tau_2\}$*:* ***sending the confidential symbols with the artificial noise seen by the legitimate receiver***. In τ_2 time slots, we send the $m\tau_2$ useful symbols represented by $\boldsymbol{v}$, superimposed by a linear combination (specified by $\boldsymbol{\Theta}$) of the artificial noise observed by the legitimate receiver in phase 1.[4]
- *Phase 3,* $t \in \mathcal{T}_3 \triangleq \{\tau_1 + \tau_2 + 1, \ldots, \tau_1 + \tau_2 + \tau_3\}$*:* ***repeating the eavesdropper's observation during phase 2***. The final phase consists in sending a linear combination (specified by $\boldsymbol{\Phi}$) of the

[4]As mentioned before, we ignore the scaling factor necessary to meet the power constraint. The same holds for the transmit vector in phase 3.

eavedropper's observation in phase 2. The aim of this phase is to complete the equations for the legitimate receiver to solve the useful symbols $\boldsymbol{v}$ without exposing anything new to the eavedropper.

After three phases, the observations are given by

$$\boldsymbol{y} = \underbrace{\begin{bmatrix} \mathbf{I}_{n_{\mathsf{A}}\tau_1} & \mathbf{0}_{m\tau_2} \\ \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta} & \tilde{\boldsymbol{H}}_2 \\ \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta} & \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2 \end{bmatrix}}_{\widehat{\mathsf{H}}} \begin{bmatrix} \tilde{\boldsymbol{H}}_1\boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} + \boldsymbol{e}, \tag{50a}$$

$$\boldsymbol{z} = \underbrace{\begin{bmatrix} \tilde{\boldsymbol{G}}_1 & \mathbf{0}_{n_{\mathsf{B}}\tau_2} \\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1 & \mathbf{I}_{n_{\mathsf{B}}\tau_2} \\ \tilde{\boldsymbol{G}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1 & \tilde{\boldsymbol{G}}_3\boldsymbol{\Phi} \end{bmatrix}}_{\widehat{\mathsf{G}}} \begin{bmatrix} \boldsymbol{u} \\ \tilde{\boldsymbol{G}}_2\boldsymbol{v} \end{bmatrix} + \boldsymbol{b}. \tag{50b}$$

Therefore, we have

$$\begin{aligned} I(\boldsymbol{v};\boldsymbol{y}) &= I(\boldsymbol{v}, \tilde{\boldsymbol{H}}_1\boldsymbol{u};\boldsymbol{y}) - I(\tilde{\boldsymbol{H}}_1\boldsymbol{u};\boldsymbol{y}\,|\,\boldsymbol{v}) && (51)\\ &\doteq \operatorname{rank}\big(\widehat{\mathsf{H}}\big)\log(P) - \operatorname{rank}\begin{pmatrix} \mathbf{I}_{n_{\mathsf{A}}\tau_1} \\ \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta} \\ \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta} \end{pmatrix}\log(P) && (52)\\ &= \left(n_{\mathsf{A}}\tau_1 + \operatorname{rank}\begin{pmatrix} \tilde{\boldsymbol{H}}_2 \\ \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2 \end{pmatrix}\right)\log(P) - n_{\mathsf{A}}\tau_1\log P && (53)\\ &= \operatorname{rank}\begin{pmatrix} \tilde{\boldsymbol{H}}_2 \\ \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2 \end{pmatrix}\log(P) && (54)\\ &= mn_{\mathsf{A}}(m-n_{\mathsf{B}})\log P, && (55) \end{aligned}$$

where (53) follows due to the block-triangular structure of $\widehat{\mathsf{H}}$ and by the fact that the rank of the second term corresponds to the rank of the identity matrix. In order to prove the last equality, we need to show first that the submatrix $\tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2$ has full-row rank with linearly independent $n_{\mathsf{A}}\tau_3$ rows. This is satisfied by letting

$$\boldsymbol{\Phi}\boldsymbol{\Pi} = \left[\operatorname{diag}(\{\boldsymbol{\Phi}_t\}_{t=1}^{\tau_3}) \quad \mathbf{0}_{m\tau_3\times(n_{\mathsf{B}}\tau_2 - n_{\mathsf{A}}\tau_3)}\right], \tag{56}$$

where $\boldsymbol{\Pi}$ is a permutation matrix such that the first $n_{\mathsf{A}}\tau_3$ rows of $\boldsymbol{\Pi}^{\mathsf{T}}\tilde{\boldsymbol{G}}_2$, is block diagonal, denoted by $\operatorname{diag}\left(\{\boldsymbol{G}_{2,t}^{\Pi}\}_{t\in\mathcal{T}_2}\right)$, where $\boldsymbol{G}_{2,t}^{\Pi} \in \mathbb{C}^{(m-n_{\mathsf{A}})\times m}$ is a submatrix of $\tilde{\boldsymbol{G}}_{2,t}$; $\boldsymbol{\Phi}_t$ denotes a $m \times n_{\mathsf{A}}$ matrix with n_{A} independent columns, e.g., $\boldsymbol{\Phi}_t = \left[\mathbf{I}_{n_{\mathsf{A}}} \quad \mathbf{0}_{n_{\mathsf{A}}\times(m-n_{\mathsf{A}})}\right]^{\mathsf{T}}$. Note that with this particular choice of $\boldsymbol{\Phi}$, the resulting submatrix is given by

$$\tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2 = \operatorname{diag}(\{\boldsymbol{H}_{t+\tau_1+\tau_2}\boldsymbol{\Phi}_t\}_{t=1}^{\tau_3})\operatorname{diag}\left(\{\boldsymbol{G}_{2,t}^{\Pi}\}_{t\in\mathcal{T}_2}\right) \tag{57}$$

Since the first matrix in the right hand side of (57) is square and full-rank, it is easy to see that $\text{rank}\begin{pmatrix}\tilde{\boldsymbol{H}}_2\\ \tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}\tilde{\boldsymbol{G}}_2\end{pmatrix} = \text{rank}\begin{pmatrix}\tilde{\boldsymbol{H}}_2\\ \text{diag}(\{\boldsymbol{G}^{\boldsymbol{\Pi}}_{2,t}\}_{t\in\mathcal{T}_2})\end{pmatrix}$. By the row permutation, we can readily show that the latter has a desired rank of $mn_{\mathsf{A}}(m-n_{\mathsf{B}})$. Namely,

$$\text{rank}\begin{pmatrix}\tilde{\boldsymbol{H}}_2\\ \text{diag}\left(\{\boldsymbol{G}^{\boldsymbol{\Pi}}_{2,t}\}_{t\in\mathcal{T}_2}\right)\end{pmatrix} = \sum_{t=1}^{n_{\mathsf{A}}(m-n_{\mathsf{B}})}\text{rank}\begin{pmatrix}\tilde{\boldsymbol{H}}_{2,t}\\ \tilde{\boldsymbol{G}}^{\boldsymbol{\Pi}}_{2,t}\end{pmatrix} \tag{58}$$

$$= n_{\mathsf{A}}(m-n_{\mathsf{B}})m, \tag{59}$$

where the last equality follows by noticing that each block t corresponds to m different rows of the state matrix $\boldsymbol{S}_t$ which are linearly independent from Assumption 2.2. On the other hand, the eavesdropper's observation is filled by the artificial noise and thus does not expose more than a vanishing fraction of the useful message, i.e.,

$$I(\boldsymbol{v};\boldsymbol{z}) \le I(\tilde{\boldsymbol{G}}_2\boldsymbol{v};\boldsymbol{z}) \tag{60}$$

$$= I(\tilde{\boldsymbol{G}}_2\boldsymbol{v},\boldsymbol{u};\boldsymbol{z}) - I(\boldsymbol{u};\boldsymbol{z}\,|\,\tilde{\boldsymbol{G}}_2\boldsymbol{v}) \tag{61}$$

$$\doteq n_{\mathsf{B}}(\tau_1+\tau_2)\log P - \text{rank}\begin{pmatrix}\tilde{\boldsymbol{G}}_1\\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\\ \tilde{\boldsymbol{G}}_3\boldsymbol{B}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\end{pmatrix}\log P \tag{62}$$

$$= mn_{\mathsf{A}}n_{\mathsf{B}}\log P - \text{rank}\begin{pmatrix}\tilde{\boldsymbol{G}}_1\\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\end{pmatrix}\log P \tag{63}$$

$$\doteq 0, \tag{64}$$

where (60) follows due to the Markov chain $\boldsymbol{v}\leftrightarrow\tilde{\boldsymbol{G}}_2\boldsymbol{v}\leftrightarrow\boldsymbol{z}$; (62) follows by noticing that the rank of $\widehat{\mathbf{G}}$ is determined by the submatrix corresponding to first two phases; (63) follows because the third block row is a linear combination of rows from the second block row. In order to prove the last equality, we choose

$$\boldsymbol{\Theta}\boldsymbol{\Pi} = \left[\text{diag}(\{\boldsymbol{\Theta}_t\}_{t=1}^{\tau_2}) \quad \mathbf{0}_{m\tau_1\times(n_{\mathsf{A}}\tau_1-n_{\mathsf{B}}\tau_2)}\right], \tag{65}$$

where $\boldsymbol{\Pi}$ is a permutation matrix[5] such that the first $n_{\mathsf{B}}\tau_2$ rows of $\boldsymbol{\Pi}^{\mathsf{T}}\tilde{\boldsymbol{H}}_1$ is block diagonal, denoted by $\text{diag}\left(\{\tilde{\boldsymbol{H}}^{\boldsymbol{\Pi}}_{1,t}\}_{t\in\mathcal{T}_1}\right)$, with $\tilde{\boldsymbol{H}}^{\boldsymbol{\Pi}}_{1,t}$ being a $(m-n_{\mathsf{B}})\times m$ submatrix of $\tilde{\boldsymbol{H}}_{1,t}$; $\boldsymbol{\Theta}_t$ denotes a $m\times n_{\mathsf{B}}$ matrix with n_{B} independent columns, e.g., $\boldsymbol{\Theta}_t = \begin{bmatrix}\mathbf{I}_{n_{\mathsf{B}}} & \mathbf{0}_{n_{\mathsf{B}}\times(m-n_{\mathsf{B}})}\end{bmatrix}^{\mathsf{T}}$. By applying exactly the same reasoning as on the choice of $\boldsymbol{\Phi}$, we can prove that $\text{rank}\begin{pmatrix}\tilde{\boldsymbol{G}}_1\\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1\end{pmatrix} = m\tau_1 = mn_{\mathsf{A}}n_{\mathsf{B}}$. As a result, the $n_{\mathsf{A}}m(m-n_{\mathsf{B}})$ useful symbols can be conveyed secretly over $n_{\mathsf{A}}n_{\mathsf{B}}+m(m-n_{\mathsf{B}})$ time slots in the high SNR regime, yielding the SDoF of $\frac{n_{\mathsf{A}}m(m-n_{\mathsf{B}})}{n_{\mathsf{A}}n_{\mathsf{B}}+m(m-n_{\mathsf{B}})}$.

[5]We abuse the notation to denote another permutation matrix than the one used in (56) .

C. Achievability proof of Theorem 2

In this subsection, we provide the achievability proof for the case of delayed partial CSIT when the transmitter has delayed CSI only on the legitimate channel. We focus on the case $\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m < n_{\mathsf{A}} + n_{\mathsf{B}}$. For the case of $m \geq n_{\mathsf{A}} + n_{\mathsf{B}}$, we can easily show that the desired SDoF follows by using only $n_{\mathsf{A}} + n_{\mathsf{B}}$ antennas out of m, i.e., by replacing m by $n_{\mathsf{A}} + n_{\mathsf{B}}$ similarly to the case of delayed CSIT on both channels. We propose a variant of the artificial noise alignment scheme described previously. The lack of CSIT on the eavesdropper channel requires the following modifications. First, the transmission consists of first two phases presented in Table II, because the lack of CSI on the eavesdropper channel does not enable the transmitter to repeat the signal overheard by the eavesdropper (corresponding to the third phase). Consequently, the confidential symbols $\boldsymbol{v}$ sent during the second phase must be decoded within this phase. This decreases the dimension of $\boldsymbol{v}$ from $m\tau_2$ to $n_{\mathsf{A}}\tau_2$. After two phases, the observations are given by

$$\boldsymbol{y} = \underbrace{\begin{bmatrix} \mathbf{I}_{n_{\mathsf{A}}\tau_1} & \mathbf{0}_{n_{\mathsf{A}}\tau_2} \\ \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta} & \tilde{\boldsymbol{H}}_2 \end{bmatrix}}_{\widehat{\mathsf{H}}} \begin{bmatrix} \tilde{\boldsymbol{H}}_1\boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} + \boldsymbol{e}, \tag{66a}$$

$$\boldsymbol{z} = \underbrace{\begin{bmatrix} \tilde{\boldsymbol{G}}_1 & \mathbf{0}_{n_{\mathsf{B}}\tau_2} \\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}\tilde{\boldsymbol{H}}_1 & \mathbf{I}_{n_{\mathsf{B}}\tau_2} \end{bmatrix}}_{\widehat{\mathsf{G}}} \begin{bmatrix} \boldsymbol{u} \\ \tilde{\boldsymbol{G}}_2\boldsymbol{v} \end{bmatrix} + \boldsymbol{b}. \tag{66b}$$

Following similar steps as before and choosing $\boldsymbol{\Theta}$ in (65), we can easily show that

$$I(\boldsymbol{v}; \boldsymbol{y}) \doteq n_{\mathsf{A}}^2(m - n_{\mathsf{B}}) \log P, \tag{67}$$

$$I(\boldsymbol{v}; \boldsymbol{z}) \doteq 0. \tag{68}$$

As a result, the $n_{\mathsf{A}}^2(m - n_{\mathsf{B}})$ useful symbols can be conveyed secretly over $n_{\mathsf{A}}m$ time slots in the high SNR regime, yielding the SDoF of $\frac{n_{\mathsf{A}}(m-n_{\mathsf{B}})}{m}$.

V. Broadcast Channel with Confidential Messages: Proof of Theorem 3

A. Converse

We focus on the case $m > \max\{n_{\mathsf{A}}, n_{\mathsf{B}}\}$ in the following. The converse for the other cases is trivial from Section IV. The secrecy constraint (5) together with Fano's inequality for W_{B}, i.e., $h(W_{\mathsf{B}}|\boldsymbol{z}^n) \leq n\epsilon$, yields

$$I(W_{\mathsf{A}}; \boldsymbol{z}^n | W_{\mathsf{B}}) \leq n\, \mathsf{o}(\log P). \tag{69}$$

Similarly to the converse of the MIMO wiretap channel, we obtain two upper bounds on R_{A}. The first bound is obtained by combining (69) with Fano's inequality on W_{A}, i.e., $h(W_{\mathsf{A}}|\boldsymbol{y}^n) \leq n\epsilon$,

$$\begin{aligned} n(R_{\mathsf{A}} - \mathsf{o}(\log P)) &\leq I(W_{\mathsf{A}}; \boldsymbol{y}^n|W_{\mathsf{B}}) - I(W_{\mathsf{A}}; \boldsymbol{z}^n|W_{\mathsf{B}}) \\ &\leq I(W_{\mathsf{A}}; \boldsymbol{y}^n|\boldsymbol{z}^n, W_{\mathsf{B}}) \qquad (70) \\ &\leq h(\boldsymbol{y}^n|\boldsymbol{z}^n, W_{\mathsf{B}}) \qquad (71) \\ &\dot{\leq} \frac{\tilde{m} - n_{\mathsf{B}}}{n_{\mathsf{B}}} h(\boldsymbol{z}^n|W_{\mathsf{B}}), \qquad (72) \end{aligned}$$

where (70) follows by $I(W_{\mathsf{A}}; \boldsymbol{y}^n|W_{\mathsf{B}}) \leq I(W_{\mathsf{A}}; \boldsymbol{y}^n, \boldsymbol{z}^n|W_{\mathsf{B}})$; (72) follows from inequality (9a) in Lemma 1. The second bound is (34) which holds also here by replacing W by W_{A}, namely,

$$I(W_{\mathsf{A}}; \boldsymbol{y}^n) - I(W_{\mathsf{A}}; \boldsymbol{z}^n) \dot{\leq} h(\boldsymbol{y}^n) - \frac{n_{\mathsf{A}}}{\tilde{m}} h(\boldsymbol{z}^n). \qquad (73)$$

Putting the two upper bounds together, we have

$$n(R_{\mathsf{A}} - \mathsf{o}(\log P)) \dot{\leq} \min\left\{ \frac{\tilde{m} - n_{\mathsf{B}}}{n_{\mathsf{B}}} h(\boldsymbol{z}^n|W_{\mathsf{B}}), h(\boldsymbol{y}^n) - \frac{n_{\mathsf{A}}}{\tilde{m}} h(\boldsymbol{z}^n) \right\}. \qquad (74)$$

On the other hand, Fano's inequality for W_{B} leads to

$$n(R_{\mathsf{B}} - \mathsf{o}(\log P)) \leq h(\boldsymbol{z}^n) - h(\boldsymbol{z}^n|W_{\mathsf{B}}). \qquad (75)$$

Now, we sum inequalities (74) and (75) with the weight $T_{\mathsf{A}} = n_{\mathsf{A}} n_{\mathsf{B}} + \tilde{m}(\tilde{m} - n_{\mathsf{B}})$, $n_{\mathsf{A}}(\tilde{m} - n_{\mathsf{B}})$, respectively. This yields

$$\begin{aligned} n(T_{\mathsf{A}} R_{\mathsf{A}} + n_{\mathsf{A}}(\tilde{m} - n_{\mathsf{B}}) R_{\mathsf{B}} - \mathsf{o}(\log P)) &\dot{\leq} \max_{h(\boldsymbol{y}^n)} \max_{\alpha} \min\left\{ (\tilde{m} - n_{\mathsf{B}})\alpha, T_{\mathsf{A}} h(\boldsymbol{y}^n) - \frac{n_{\mathsf{A}} n_{\mathsf{B}}}{\tilde{m}} \alpha \right\} \qquad (76) \\ &\leq \max_{h(\boldsymbol{y}^n)} \tilde{m}(\tilde{m} - n_{\mathsf{B}}) h(\boldsymbol{y}^n) \qquad (77) \\ &\dot{\leq} n_{\mathsf{A}} \tilde{m}(\tilde{m} - n_{\mathsf{B}}) n \log P, \qquad (78) \end{aligned}$$

where we let $\alpha = n_{\mathsf{A}} h(\boldsymbol{z}^n) + \frac{\tilde{m}(\tilde{m} - n_{\mathsf{B}})}{n_{\mathsf{B}}} h(\boldsymbol{z}^n|W_{\mathsf{B}})$ in the first inequality and the last inequality follows because $h(\boldsymbol{y}^n) \leq n n_{\mathsf{A}} \log P + \mathsf{o}(\log P)$. By dividing both sides by $n_{\mathsf{A}} \tilde{m}(\tilde{m} - n_{\mathsf{B}}) \log P$ and letting P grow, we obtain the first desired inequality (15a). Due to the symmetry of the problem, (15b) can be obtained by swapping the roles of R_{A} and R_{B}. This completes the converse proof.

B. Achievability

The corner points can be achieved by the ANA scheme described in Section IV. Here, we provide a strategy achieving the sum SDoF point. In fact, the ANA scheme for the MIMO wiretap channel

TABLE IV

PROPOSED FOUR-PHASE SCHEME FOR $\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$.

Phase 1	Phase 2	Phase 3
$\boldsymbol{x}_1 = \boldsymbol{u}$	$\boldsymbol{x}_2 = \boldsymbol{v}_{\mathsf{A}} + \boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{y}}_1$	$\boldsymbol{x}_3 = \boldsymbol{v}_{\mathsf{B}} + \boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{z}}_1$
$\tilde{\boldsymbol{y}}_1 = \tilde{\boldsymbol{H}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{y}}_2 = \tilde{\boldsymbol{H}}_2\left(\boldsymbol{v}_{\mathsf{A}} + \boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u}\right)$	$\tilde{\boldsymbol{y}}_3 = \tilde{\boldsymbol{H}}_3\left(\boldsymbol{v}_{\mathsf{B}} + \boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u}\right)$
$\tilde{\boldsymbol{z}}_1 = \tilde{\boldsymbol{G}}_1\boldsymbol{u}$	$\tilde{\boldsymbol{z}}_2 = \tilde{\boldsymbol{G}}_2\left(\boldsymbol{v}_{\mathsf{A}} + \boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u}\right)$	$\tilde{\boldsymbol{z}}_3 = \tilde{\boldsymbol{G}}_3\left(\boldsymbol{v}_{\mathsf{B}} + \boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u}\right)$

Phase 4
$\boldsymbol{x}_4 = \boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{z}}_2 + \boldsymbol{\Phi}_{\mathsf{B}}\tilde{\boldsymbol{y}}_3$
$\tilde{\boldsymbol{y}}_4 = \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{A}}(\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u}) + \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{B}}(\tilde{\boldsymbol{H}}_3\boldsymbol{v}_{\mathsf{B}} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u})$
$\tilde{\boldsymbol{z}}_4 = \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{A}}(\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u}) + \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{B}}(\tilde{\boldsymbol{H}}_3\boldsymbol{v}_{\mathsf{B}} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u})$

TABLE V

LENGTH OF FOUR PHASES $\{\tau_i\}$ FOR DIFFERENT m, n_{A}, AND n_{B}.

	$\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$	$m > n_{\mathsf{A}} + n_{\mathsf{B}}$
τ_1	$n_{\mathsf{A}}n_{\mathsf{B}}$	$n_{\mathsf{A}}n_{\mathsf{B}}$
τ_2	$n_{\mathsf{A}}(m - n_{\mathsf{B}})$	n_{A}^2
τ_3	$n_{\mathsf{B}}(m - n_{\mathsf{A}})$	n_{B}^2
τ_4	$(m - n_{\mathsf{A}})(m - n_{\mathsf{B}})$	$n_{\mathsf{A}}n_{\mathsf{B}}$
total duration $\sum_{i=1}^4 \tau_i$	m^2	$(n_{\mathsf{A}} + n_{\mathsf{B}})^2$

in Section IV can be suitably modified to convey two confidential messages. We focus on the case $\max\{n_{\mathsf{A}}, n_{\mathsf{B}}\} < m \leq n_{\mathsf{A}} + n_{\mathsf{B}}$ because the converse proof says that we only need to use $n_{\mathsf{A}} + n_{\mathsf{B}}$ antennas for the case $m > n_{\mathsf{A}} + n_{\mathsf{B}}$.

The proposed four-phase scheme is presented in Table IV, where the signal model is describe concisely with the block matrix notation:

$$\boldsymbol{u} = \left[\begin{array}{ccc} \boldsymbol{u}_1^{\mathsf{T}} & \ldots & \boldsymbol{u}_{\tau_1}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}} \in \mathbb{C}^{m\tau_1 \times 1}, \tag{79}$$

$$\boldsymbol{v}_{\mathsf{A}} = \left[\begin{array}{ccc} \boldsymbol{v}_{\mathsf{A},1}^{\mathsf{T}} & \ldots & \boldsymbol{v}_{\mathsf{A},\tau_2}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}} \in \mathbb{C}^{m\tau_2 \times 1}, \qquad \boldsymbol{v}_{\mathsf{B}} = \left[\begin{array}{ccc} \boldsymbol{v}_{\mathsf{B},1}^{\mathsf{T}} & \ldots & \boldsymbol{v}_{\mathsf{B},\tau_3}^{\mathsf{T}} \end{array}\right]^{\mathsf{T}} \in \mathbb{C}^{m\tau_3 \times 1}, \tag{80}$$

$$\tilde{\boldsymbol{H}}_i = \mathrm{diag}\left(\{\boldsymbol{H}_t\}_{t\in\mathcal{T}_i}\right) \in \mathbb{C}^{n_{\mathsf{A}}\tau_i \times m\tau_i}, \qquad \tilde{\boldsymbol{G}}_i = \mathrm{diag}\left(\{\boldsymbol{G}_t\}_{t\in\mathcal{T}_i}\right) \in \mathbb{C}^{n_{\mathsf{B}}\tau_i \times m\tau_i}, \tag{81}$$

$$\boldsymbol{\Theta}_{\mathsf{A}} \in \mathbb{C}^{m\tau_2 \times n_{\mathsf{A}}\tau_1}, \qquad \boldsymbol{\Theta}_{\mathsf{B}} \in \mathbb{C}^{m\tau_3 \times n_{\mathsf{B}}\tau_1}, \tag{82}$$

$$\boldsymbol{\Phi}_{\mathsf{A}} \in \mathbb{C}^{m\tau_4 \times n_{\mathsf{B}}\tau_2}, \qquad \boldsymbol{\Phi}_{\mathsf{B}} \in \mathbb{C}^{m\tau_4 \times n_{\mathsf{A}}\tau_3}, \tag{83}$$

where the durations of four phases $\{\tau_i\}_i$ are given in Table V. The four phases consist of:

- *Phase 1, $t \in \mathcal{T}_1 \triangleq \{1, \ldots, \tau_1\}$: **sending the artificial noise***. The $m\tau_1$ symbols sent in τ_1 time slots

is represented by $\boldsymbol{u}$.

- *Phase 2,* $t \in \mathcal{T}_2 \triangleq \{\tau_1+1,\ldots,\tau_1+\tau_2\}$*:* ***sending the confidential symbols*** $\boldsymbol{v_{\mathsf{A}}}$ ***with the artificial noise seen by receiver*** $\boldsymbol{\mathsf{A}}$. In τ_2 time slots, we send the $m\tau_2$ useful symbols represented by $\boldsymbol{v}_{\mathsf{A}}$, superimposed by a linear combination (specified by $\boldsymbol{\Theta}_{\mathsf{A}}$) of the artificial noise observed by receiver A in phase 1.
- *Phase 3,* $t \in \mathcal{T}_3 \triangleq \{\tau_1+\tau_2+1,\ldots,\tau_1+\tau_2+\tau_3\}$*:* ***sending the confidential symbols*** $\boldsymbol{v_{\mathsf{B}}}$ ***with the artificial noise seen by receiver*** $\boldsymbol{\mathsf{B}}$. In τ_3 time slots, we send the $m\tau_3$ useful symbols represented by $\boldsymbol{v}_{\mathsf{B}}$, superimposed by a linear combination (specified by $\boldsymbol{\Theta}_{\mathsf{B}}$) of the artificial noise observed by receiver B in phase 1.
- *Phase 4,* $t \in \mathcal{T}_4 \triangleq \{\tau_1+\tau_2+\tau_3+1,\ldots,\tau_1+\tau_2+\tau_3+\tau_4\}$*:* ***repeating the past observations during in phase 2 and 3***. The final phase consists in sending a linear combination of receiver B's observation in phase 2 (specified by $\boldsymbol{\Phi}_{\mathsf{A}}$) and receiver A's observation in phase 3 (specified by $\boldsymbol{\Phi}_{\mathsf{B}}$). The aim of this phase is to complete the equations for the intended receivers to solve the useful symbols without exposing anything new about the message to the unintended receivers.

After four phases, the observations are given by

$$\boldsymbol{y} = \underbrace{\begin{bmatrix} \tilde{\boldsymbol{H}}_2 & \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta}_{\mathsf{A}} & \mathbf{0} \\ \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2 & \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}} & \tilde{\boldsymbol{H}}_4 \\ \mathbf{0} & \mathbf{I}_{n_{\mathsf{A}}\tau_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_{\mathsf{A}}\tau_3} \end{bmatrix}}_{\widehat{\mathsf{H}}_{\text{bcc}}} \begin{bmatrix} \boldsymbol{v}_{\mathsf{A}} \\ \tilde{\boldsymbol{H}}_1\boldsymbol{u} \\ \tilde{\boldsymbol{H}}_3\boldsymbol{v}_{\mathsf{B}} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u} \end{bmatrix} + \boldsymbol{e}, \tag{84a}$$

$$\boldsymbol{z} = \underbrace{\begin{bmatrix} \mathbf{0} & \mathbf{I}_{n_{\mathsf{B}}\tau_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_{\mathsf{B}}\tau_2} \\ \tilde{\boldsymbol{G}}_3 & \tilde{\boldsymbol{G}}_3\boldsymbol{\Theta}_{\mathsf{B}} & \mathbf{0} \\ \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{B}}\tilde{\boldsymbol{H}}_3 & \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{B}}\tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}} & \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{A}} \end{bmatrix}}_{\widehat{\mathsf{G}}_{\text{bcc}}} \begin{bmatrix} \boldsymbol{v}_{\mathsf{B}} \\ \tilde{\boldsymbol{G}}_1\boldsymbol{u} \\ \tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u} \end{bmatrix} + \boldsymbol{b}. \tag{84b}$$

First, we examine the mutual information between $\boldsymbol{v}_{\mathsf{A}}$ and $\boldsymbol{y}$:

$$I(\boldsymbol{v}_{\mathsf{A}};\boldsymbol{y}) = I(\boldsymbol{v}_{\mathsf{A}}, \tilde{\boldsymbol{H}}_1\boldsymbol{u}, \tilde{\boldsymbol{H}}_3\boldsymbol{v}_{\mathsf{B}} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u};\boldsymbol{y}) - I(\tilde{\boldsymbol{H}}_1\boldsymbol{u}, \tilde{\boldsymbol{H}}_3\boldsymbol{v}_{\mathsf{B}} + \tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1\boldsymbol{u};\boldsymbol{y}\,|\,\boldsymbol{v}_{\mathsf{A}}) \tag{85}$$

$$\doteq \text{rank}(\widehat{\mathbf{H}}_{\text{bcc}})\log(P) - \text{rank}\begin{pmatrix} \tilde{\boldsymbol{H}}_2\boldsymbol{\Theta}_{\mathsf{A}} & \mathbf{0} \\ \tilde{\boldsymbol{H}}_4\tilde{\boldsymbol{H}}_3\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}} & \tilde{\boldsymbol{H}}_4 \\ \mathbf{I}_{n_{\mathsf{A}}\tau_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_{\mathsf{A}}\tau_3} \end{pmatrix}\log P \tag{86}$$

$$= \left(n_{\mathsf{A}}(\tau_1+\tau_3) + \text{rank}\begin{pmatrix} \tilde{\boldsymbol{H}}_2 \\ \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2 \end{pmatrix}\right)\log P - n_{\mathsf{A}}(\tau_1+\tau_3)\log P \tag{87}$$

$$= \text{rank}\begin{pmatrix} \tilde{\boldsymbol{H}}_2 \\ \tilde{\boldsymbol{H}}_4\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2 \end{pmatrix}\log P \tag{88}$$

$$= m\,n_{\mathsf{A}}(m-n_{\mathsf{B}})\log P, \tag{89}$$

where in (87) the first term is due to the block-triangular structure of $\widehat{\mathbf{H}}_{\text{bcc}}$ and the second term follows because the rank corresponds to the rank of the identity matrix; (89) follows by choosing $\boldsymbol{\Phi}_{\mathsf{A}}$ given in (56) where we replace τ_3 with τ_4.

Next, in order to examine the leakage of $\boldsymbol{v}_{\mathsf{A}}$ to receiver B, we write

$$I(\boldsymbol{v}_{\mathsf{A}};\boldsymbol{z},\boldsymbol{v}_{\mathsf{B}}) = I(\boldsymbol{v}_{\mathsf{A}};\boldsymbol{z}|\boldsymbol{v}_{\mathsf{B}}) \tag{90}$$

$$\leq I(\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}};\boldsymbol{z}|\boldsymbol{v}_{\mathsf{B}}) \tag{91}$$

$$= I(\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}},\boldsymbol{u};\boldsymbol{z}|\boldsymbol{v}_{\mathsf{B}}) - I(\boldsymbol{u};\boldsymbol{z}\,|\,\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}},\boldsymbol{v}_{\mathsf{B}}) \tag{92}$$

$$\leq I(\tilde{\boldsymbol{G}}_1\boldsymbol{u},\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u};\boldsymbol{z}|\boldsymbol{v}_{\mathsf{B}}) - I(\boldsymbol{u};\boldsymbol{z}\,|\,\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}},\boldsymbol{v}_{\mathsf{B}}) \tag{93}$$

$$\doteq \text{rank}\begin{pmatrix} \mathbf{I}_{n_{\mathsf{B}}\tau_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_{\mathsf{B}}\tau_2} \\ \tilde{\boldsymbol{G}}_3\boldsymbol{\Theta}_{\mathsf{B}} & \mathbf{0} \\ \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{B}}\tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}} & \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{A}} \end{pmatrix}\log P - \text{rank}\begin{pmatrix} \tilde{\boldsymbol{G}}_1 \\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1 \\ \tilde{\boldsymbol{G}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1 \\ \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{B}}\tilde{\boldsymbol{H}}_3\boldsymbol{\Theta}_{\mathsf{B}}\tilde{\boldsymbol{G}}_1 + \tilde{\boldsymbol{G}}_4\boldsymbol{\Phi}_{\mathsf{A}}\tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1 \end{pmatrix}\log P \tag{94}$$

$$= n_{\mathsf{B}}(\tau_1+\tau_2)\log P - \text{rank}\begin{pmatrix} \tilde{\boldsymbol{G}}_1 \\ \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1 \end{pmatrix}\log P \tag{95}$$

$$\doteq 0, \tag{96}$$

where (91) follows from the Markov chain $\boldsymbol{v}_{\mathsf{A}} \leftrightarrow \tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} \leftrightarrow \boldsymbol{z}$; (92) is due to another Markov chain $(\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}},\boldsymbol{u}) \leftrightarrow (\tilde{\boldsymbol{G}}_1\boldsymbol{u},\tilde{\boldsymbol{G}}_2\boldsymbol{v}_{\mathsf{A}} + \tilde{\boldsymbol{G}}_2\boldsymbol{\Theta}_{\mathsf{A}}\tilde{\boldsymbol{H}}_1\boldsymbol{u}) \leftrightarrow \boldsymbol{z}$; in (95) we notice that two block columns of $\widehat{\mathbf{G}}_{\text{bcc}}$ is block-triangular and the second term follows by keeping only linearly independent block rows; the last equality is obtained by setting $\boldsymbol{\Theta}_{\mathsf{A}}$ given in (65). As a result, the SDoF $d_{\mathsf{A}} = \frac{n_{\mathsf{A}}(\tilde{m}-n_{\mathsf{B}})}{\tilde{m}}$ is achieved with the proposed scheme. By symmetry of the problem, we have $d_{\mathsf{B}} = \frac{n_{\mathsf{B}}(\tilde{m}-n_{\mathsf{A}})}{\tilde{m}}$ which completes the proof.

VI. Conclusions and Perspectives

We studied the impact of delayed CSIT on the MIMO wiretap channel and the MIMO broadcast channel with confidential messages (BCC) by focusing on the secrecy degrees of freedom (SDoF) metric. The

optimal SDoF region of the two-user Gaussian MIMO-BCC was fully characterized. It is shown that an artificial noise alignment (ANA) scheme, which can be regarded as a non-trivial extension of Maddah-Ali Tse (MAT) scheme, can achieve the entire SDoF region. The proposed ANA scheme enables to nicely quantify the resource overhead to be dedicated to secure the confidential messages, which in turn appears as a DoF loss. Although delayed CSIT was found useful to improve the SDoF over a wide range of the MIMO system, our study somehow revealed the bottleneck of physical-layer security due to its high sensitivity to the quality of CSIT.

Several interesting open problems emerge out of this work. First, some techniques used for lower- and upper-bounding the SDoF in this work may serve to enhance further insights on related problems for moderate SNR regimes. Second, the characterization of the SDoF upper bound of the Gaussian MIMO wiretap channel with delayed partial CSIT remains open. We emphasize that for the case of partial CSIT, the inequalities due to the channel symmetry still hold true, but these do not seem to be enough to prove the converse. The challenge consists of finding novel and tighter inequalities that capture some new asymmetry between $h(\boldsymbol{z}^n)$ and $h(\boldsymbol{y}^n)$. Finally, the extension to more complex scenarios such as the BCC with more than two receivers can be also investigated.

APPENDIX A
PROOF OF LEMMA 1

Lemma 3: Let $x^L = (x_1, \ldots, x_L)$ be entropy-symmetric such that $h(\{x_j : j \in \mathcal{J}\}) = h(\{x_k : k \in \mathcal{K}\})$, for any $|\mathcal{J}| = |\mathcal{K}| \leq L$. Then, for any $M \geq N$, we have

$$h(x^{N+k}) - h(x^N) \geq h(x^{M+k}) - h(x^M), \quad \forall\, k \geq 0, \tag{97a}$$

$$\text{and} \quad M\, h(x^N) \geq N\, h(x^M), \tag{97b}$$

where we define $h(\emptyset) = 0$ for convenience of notation.

Proof: For $M = N$, the inequalities (97a) and (97b) hold with equality trivially. It is thus without loss of generality to assume that $M > N$.

We first prove inequality (97a). It is readily shown that

$$h(x^{N+k}) - h(x^N) = h(x_1, \ldots, x_{N+k}) - h(x_1, \ldots, x_N) \tag{98}$$

$$= h(x_1, \ldots, x_{N+k}) - h(x_{k+1}, \ldots, x_{N+k}) \tag{99}$$

$$= h(x_1, \ldots, x_k \,|\, x_{k+1}, \ldots, x_{N+k}), \tag{100}$$

where (99) is from the entropy-symmetry of x^L. Since the last equality is decreasing with $N \geq 0$, (97a) is immediate.

For the inequality (97b), we prove it by induction on L. For $L = 2$, the only non-trivial case is $M = 2$ and $N = 1$, where we have

$$2h(x_1) = h(x_1) + h(x_2) \tag{101}$$

$$\geq h(x_1, x_2). \tag{102}$$

Assume that the result holds to $L = l - 1$, i.e., (97b) is true for any $(M, N) \in \{(j, k) : \ l - 1 \geq j > k\}$. We would like to prove that it holds for any $(M, N) \in \{(j, k) : \ l \geq j > k\}$. In particular, all we need to prove is that the inequality holds for $M = l$ and any $N \leq l - 1$, i.e.,

$$l\, h(x^N) - N\, h(x^l) \geq 0, \quad \forall\, N \leq l - 1. \tag{103}$$

To this end, we first write

$$l\, h(x^N) - N\, h(x^l) = (l - N)\, h(x^N) - N\, (h(x^l) - h(x^N)). \tag{104}$$

For N such that $l - N \leq N$, we can lower-bound the right hand side (RHS) of (104) as

$$(l - N)h(x^N) - N(h(x^l) - h(x^N)) \geq (l - N)h(x^N) - N(h(x^N) - h(x^{2N-l})) \tag{105}$$

$$= N\, h(x^{2N-l}) - (2N - l)h(x^N) \tag{106}$$

$$\geq 0, \tag{107}$$

where the first inequality is from the fact that applying (97a), $h(x^l) - h(x^N) \leq h(x^N) - h(x^{2N-l})$; the last inequality is from the induction assumption, since $(N, 2N - l)$ is such that $l - 1 \geq N \geq 2N - l$.

For N such that $l - N > N$, we lower-bound the RHS of (104) differently

$$(l - N)h(x^N) - N(h(x^l) - h(x^N)) \geq (l - N)h(x^N) - N\, h(x^{l-N}) \tag{108}$$

$$\geq 0, \tag{109}$$

where the first inequality is from the fact that applying (97a), $h(x^l) - h(x^N) \leq h(x^{l-N}) - h(x^0)$ with $h(x^0) = h(\emptyset) = 0$ by definition; the last inequality is from the induction assumption since $(l - N, N)$ is such that $l - 1 \geq l - N > N$. The proof for (103) in complete.

■

By symmetry of the problem, we only need to prove (9a). We first consider the case $n_{\mathsf{A}} + n_{\mathsf{B}} \leq m$.

$$n_{\mathsf{B}} h(\boldsymbol{y}^n, \boldsymbol{z}^n) = n_{\mathsf{B}} \sum_{t=1}^{n} h(\boldsymbol{y}_t, \boldsymbol{z}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \tag{110}$$

$$= n_{\mathsf{B}} \sum_{t=1}^{n} h(\boldsymbol{\omega}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \tag{111}$$

$$\leq (n_{\mathsf{A}} + n_{\mathsf{B}}) \sum_{t=1}^{n} h(\boldsymbol{z}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \tag{112}$$

$$\leq (n_{\mathsf{A}} + n_{\mathsf{B}}) \sum_{t=1}^{n} h(\boldsymbol{z}_t \,|\, \boldsymbol{z}^{t-1}) \tag{113}$$

$$\leq (n_{\mathsf{A}} + n_{\mathsf{B}}) h(\boldsymbol{z}^n), \tag{114}$$

where we define $\boldsymbol{\omega}_t = \{\boldsymbol{y}_t, \boldsymbol{z}_t\}$; (112) is the application of (9a).

When $m < n_{\mathsf{A}} + n_{\mathsf{B}}$, (112) is loose. We tighten the bound as follows.

$$n_{\mathsf{B}} h(\boldsymbol{y}^n, \boldsymbol{z}^n) = n_{\mathsf{B}} \sum_{t=1}^{n} h(\boldsymbol{\omega}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \tag{115}$$

$$= n_{\mathsf{B}} \sum_{t=1}^{n} \big(h(\bar{\boldsymbol{\omega}}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) + h(\hat{\boldsymbol{\omega}}_t \,|\, \bar{\boldsymbol{\omega}}_t, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \big) \tag{116}$$

$$\leq n_{\mathsf{B}} \sum_{t=1}^{n} h(\bar{\boldsymbol{\omega}}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) + \mathsf{o}(\log P) \tag{117}$$

$$\leq m \sum_{t=1}^{n} h(\boldsymbol{z}_t \,|\, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) + \mathsf{o}(\log P) \tag{118}$$

$$\leq m \sum_{t=1}^{n} h(\boldsymbol{z}_t \,|\, \boldsymbol{z}^{t-1}) + \mathsf{o}(\log P) \tag{119}$$

$$\leq m h(\boldsymbol{z}^n) + \mathsf{o}(\log P), \tag{120}$$

where we partition $\boldsymbol{\omega}_t$ as $\boldsymbol{\omega}_t = \{\bar{\boldsymbol{\omega}}_t,\ \hat{\boldsymbol{\omega}}_t\}$ in such a way that $\bar{\boldsymbol{\omega}}_t$ and $\tilde{\boldsymbol{\omega}}_t$ are of length m and $n_{\mathsf{A}} + n_{\mathsf{B}} - m$, respectively; (117) is from the fact that $h(\hat{\boldsymbol{\omega}}_t \,|\, \bar{\boldsymbol{\omega}}_t, \boldsymbol{y}^{t-1}, \boldsymbol{z}^{t-1}) \leq h(\hat{\boldsymbol{\omega}}_t \,|\, \bar{\boldsymbol{\omega}}_t)$ and that $h(\hat{\boldsymbol{\omega}}_t \,|\, \bar{\boldsymbol{\omega}}_t) \leq \mathsf{o}(\log P)$ with the same reasoning applied in (23)-(27).

REFERENCES

[1] M.A. Maddah-Ali and D. Tse, "On the Degrees of Freedom of MISO Broadcast Channels with Delayed Feedback," *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2010-122, Sep*, pp. 2010–122, 2010.

[2] R. Liu, T. Liu, H.V. Poor, and S. Shamai (Shitz), "Multiple-Input Multiple-Output Gaussian Broadcast Channels with Confidential Messages," *IEEE Trans. on Inform. Theory*, vol. 56, no. 9, pp. 4215–4227, 2010.

[3] T. Liu and S. Shamai (Shitz), "A Note on the Secrecy Capacity of the Multiple-Antenna Wiretap Channel," *IEEE Trans. on Inform. Theory*, vol. 55, no. 6, pp. 2547–2553, 2009.

[4] A. Khisti and G.W. Wornell, "Secure Transmission with Multiple Antennas I: The MISOME Wiretap Channel," *IEEE Trans. on Inform. Theory*, vol. 56, no. 7, pp. 3088–3104, 2010.

[5] A. Khisti and G.W. Wornell, "Secure Transmission with Multiple Antennas–Part II: The MIMOME Wiretap Channel," *IEEE Trans. on Inform. Theory*, vol. 56, no. 11, pp. 5515–5532, 2010.

[6] F. Oggier and B. Hassibi, "The Secrecy Capacity of the MIMO Wiretap Channel," *IEEE Trans. on Inform. Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[7] R. Bustin, R. Liu, H. V. Poor, and S. Shamai (Shitz), "An MMSE Approach to the Secrecy Capacity of the MIMO Gaussian Wiretap Channel," *EURASIP Journal on Wireless Communications and Networking, special issue on Wireless Physical Security*, 2009.

[8] Y. Liang and G. Kramer and H. V. Poor and S. Shamai (Shitz), "Compound Wiretap Channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, 2009.

[9] A. Khisti, "Interference Alignment for the Multiantenna Compound Wiretap Channel," *IEEE Trans. on Inform. Theory*, vol. 57, no. 5, pp. 2976–2993, 2011.

[10] M. Kobayashi, Y. Liang, S. Shamai (Shitz), and M. Debbah, "On the Compound MIMO Broadcast Channels with Confidential Messages," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'09), Seoul, Korea.*, 2009, pp. 1283–1287.

[11] M. Kobayashi, J. Piantanida, S. Yang, and S. Shamai (Shitz), "On the Secrecy Degrees of Freedom of the Multi-Antenna Block Fading Wiretap Channels," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 703–711, Sept. 2011.

[12] C.S. Vaze and M.K. Varanasi, "The Degrees of Freedom Region of the Two-User MIMO Broadcast Channel with Delayed CSI," *Arxiv preprint arXiv:1101.0306*, 2010.

[13] H. Maleki, S.A. Jafar, and S. Shamai (Shitz), "Retrospective Interference Alignment," *Arxiv preprint arXiv:1009.3593*, 2010.

[14] C.S. Vaze and M.K. Varanasi, "The Degrees of Freedom Region and Interference Alignment for the MIMO Interference Channel with Delayed CSI," *Arxiv preprint arXiv:1101.5809*, 2011.

[15] M.J. Abdoli, A. Ghasemi, and A.K. Khandani, "On the Degrees of Freedom of K-User SISO Interference and X Channels with Delayed CSIT," *Arxiv preprint arXiv:1109.4314*, 2011.

[16] S. A. Jafar, "Interference Alignment: A New Look at Signal Dimensions in a Communication Network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, pp. 1–136, 2011.